Exp Econ
DOI 10.1007/s10683-017-9522-7

ORIGINAL PAPER

Why did he do that? Using counterfactuals to study the effect of intentions in extensive form games

Yola Engler[1] · **Rudolf Kerschbamer**[2] · **Lionel Page**[1]

Received: 22 March 2016 / Revised: 21 February 2017 / Accepted: 27 February 2017
© Economic Science Association 2017

Abstract We investigate the role of intentions in two-player two-stage games. For this purpose we systematically vary the set of opportunity sets the first mover can choose from and study how the second mover reacts not only to opportunities of gains but also of losses created by the choice of the first mover. We find that the possibility of gains for the second mover (generosity) and the risk of losses for the first mover (vulnerability) are important drivers for second mover behavior. On the other hand, efficiency concerns and an aversion against violating trust seem to be far less important motivations. We also find that second movers compare the actual choice of the first mover and the alternative choices that would have been available to him to allocations that involve equal material payoffs.

Keywords Social preferences · Other-regarding preferences · Intentions · Reciprocity · Trust game · Experimental economics · Behavioral economics

JEL Classification C91 · C92 · D63 · D64

Electronic supplementary material The online version of this article (doi:10.1007/s10683-017-9522-7) contains supplementary material, which is available to authorized users.

✉ Lionel Page
lionel.page@qut.edu.au

Yola Engler
yolaengler@gmail.com

Rudolf Kerschbamer
rudolf.kerschbamer@uibk.ac.at

[1] School of Economics and Finance, Queensland University of Technology, 2 George Street, Brisbane, Australia

[2] Department of Economics, University of Innsbruck, Universitaetsstrasse 15, Innsbruck, Australia

Published online: 16 March 2017

1 Introduction

Other-regarding preferences capture people's valuation not only for their own material resources but also for the material payoffs of other individuals as well as the perceived kindness of others' behavior. The theoretical literature on such preferences can be divided into two broad classes: models with distributional (unconditional) other-regarding preferences and models with intention- or action-based (conditional) other-regarding preferences.

The distributional (or "social") preference approach focuses on preferences over allocations of resources which are driven by distributional properties of the allocations. The altruism models by Andreoni and Miller (2002) and by Cox and Sadiraj (2007) fall into this category, as well as the models of inequality-aversion by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000), and the model of altruism and spite by Levine (1998).[1]

The conditional other-regarding preference approach, on the other hand, tries to explain findings neither consistent with self-regarding preferences nor in line with existing models of distributional concerns by agents' desire to react to others' intentions or actions. In this strand, a second mover's preferences in a two-person two-stage game typically become more or less benevolent depending on the perceived "kindness" of the first mover, and kindness is typically interpreted as generosity.

Two approaches have been proposed to investigate conditional other-regarding preferences theoretically. First, in psychological game theory, a player evaluates another person's kindness by forming beliefs on what the other person believes the consequences of his choice are (see Rabin 1993; Dufwenberg and Kirchsteiger 2004, for instance). This necessarily involves second-order beliefs entering the picture. Models incorporating second-order beliefs provide quite sophisticated theories of conditional other-regarding preferences. Unfortunately, they often yield multiple equilibria even in quite simple games and finding these is often not trivial. Also, empirical tests of models in this class must deal with the important and non-trivial question on how to elicit or induce second-order beliefs in a clean way.

To avoid these problems, a second approach of modeling conditional other-regarding preferences—the "revealed intentions" approach—has been proposed by Cox et al. (2007, 2008b). In this approach a second mover's benevolence in a two-player two-stage game is a function of the relative kindness or unkindness of the first mover as revealed by the objective characteristics of his (observed) choices. The first mover's kindness, in turn, is determined by the relative generosity of the opportunity set implied by his choice relative to alternative opportunity sets he could have chosen instead.

The present paper contributes to the revealed intentions approach of conditional other-regarding preferences by exposing subjects in the lab to a large number of two-player two-stage games and by studying how second movers react to the opportunities of gains and losses for each player generated by the choice of the first

[1] Another example for a model where decisions are shaped by distributional properties of the available allocations is the quasi-maximin model by Charness and Rabin (2002), which adds to material self-interest surplus maximization and the Rawlsian maximin motive as drivers for behavior.

mover. Specifically, we expose subjects in the lab to graphical representations of two-player two-stage games in which (1) the first mover has to choose between two budget sets, one containing a single allocation, the other containing several possible payoff allocations; and (2) the second mover has to choose one of the available payoff allocations in the non-trivial budget set—provided the first mover has chosen it. By systematically varying the two budget sets available to the first mover, we investigate how opportunities of gains and losses for each player influence the second mover's benevolence towards the first mover.

We find that the possibility of gains for the second mover (generosity) and the risk of losses for the first mover (vulnerability) are important drivers for second mover behavior. On the other hand, efficiency concerns and an aversion against violating trust seem to be far less important motivators. We also find that second movers compare the actual choice of the first mover and the alternative choices that would have been available to him to allocations that involve equal material payoffs.

Compared to the existing literature on conditional other-regarding preferences the present paper makes three critical contributions: The first contribution is the introduction and implementation of an experimental design in which subjects are exposed to geometric representations of choice sets. This allows for the collection of a large number of observations per subject which facilitates statistical analysis at the level of the individual decision maker. Regarding this contribution the paper closest to ours is Fisman et al. (2007). Those authors are interested in *unconditional* other-regarding concerns. As a consequence, in their experiments there is only one player role—that of a dictator—and each dictator is exposed to 50 different decision problems, each graphically represented as a linear budget set from which the subject can choose.[2] Since our main research focus is on *conditional* other-regarding preferences we extend this approach by having two player roles—the role of a first mover and the role of a second mover; the first mover chooses among graphical representations of opportunity sets while the second mover makes a dictator decision within a given opportunity set similar to the one subjects are asked to make in Fisman et al. (2007). By varying the set of budget sets available to the first mover we are able to investigate how the second mover's choice varies with the budget set actually chosen by the first mover and with the counterfactual alternative opportunity set the first mover could have chosen instead.

Our second innovation is the experimental investigation of the relative importance of different motives for behavior of players in extensive-form games. In this respect the papers closest to ours are Cox (2004) and Cox et al. (2007, 2008b, 2016). While Cox (2004) employs a triadic experimental design to disentangle the relative importance of conditional and unconditional other-regarding preferences for behavior of second movers in the investment game, the present paper's main aim is to disentangle the relative importance of different basic motives for the conditional part

[2] This is the baseline experiment in Fisman et al. (2007). In addition to this the authors also investigate two alternative treatments: one has linear budget sets as the baseline but differs from the latter in that each dictator decision has now consequences for two other persons (i.e., budget sets are three-dimensional in this treatment); the other has two-dimensional budget sets as the baseline but differs from the latter in having allocations in the choice set that differ only in the material payoff of the recipient, or only in the material payoff of the dictator (i.e., budgets are step-shaped in this treatment).

of players' other-regarding preferences. Similar to Cox et al. (2007, 2008b) we suppose that the second mover in a two-player two-stage game cares about how the opportunity set chosen by the first mover compares to alternative opportunity sets the first mover could have chosen instead. However, while these papers compare opportunity sets in terms of generosity by the first mover towards the second mover and focus on reciprocity as possible motivation for the second mover, we look not only at the possible gains for both players but also at possible losses and look at a broader array of possible motivations. In this latter respect our paper is similar to Cox et al. (2016). However, in contrast to that work we look not only on trust game constellations and we also collect many observations per individual.[3] The latter feature of our experimental design allows us to estimate utility functions at the individual level in a within-subjects design while Cox et al. (2016) derive their results from comparisons of aggregate data across treatments in a between-subjects design.

Our third innovation is the introduction of a silent social norm—the equal-split norm—into the revealed intentions approach. In this respect our paper is related to previous work on the importance of the equality norm for economic behavior—see Fehr and Schmidt (1999), Bolton and Ockenfels (2000) and Andreoni and Bernheim (2009), for instance. While Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) stress the importance of the equal-split norm for unconditional other-regarding preferences, we show that this norm is also crucial for our understanding of conditional other-regarding preferences. Conditional other-regarding preferences might also be relevant for behavior in the experiments reported by Andreoni and Bernheim (2009). However, while Andreoni and Bernheim are interested in the impact of "audience effects" on behavior, we are interested in situations where audience effects are unlikely to play a role.

The remainder of the paper is organized as follows: Sect. 2 presents our experimental design. It is followed by our conceptual framework in Sect. 3, which consists of a classification of choice characteristics, our model of social preferences, and predictions derived from the model. In Sect. 4, we report our data and estimate the parameters of our model. Section 5 discusses our findings and concludes.

2 Experimental design

Our workhorse is a two-stage game with two players. In the first stage, the first mover (FM, he) makes a binary decision—he chooses between a fixed allocation (consisting of a payoff for himself and a payoff for the second mover) and an opportunity set containing several possible allocations. In the second stage, the second mover (SM, she) chooses a fixed allocation from the opportunity set whenever the FM has chosen this option—otherwise she has no move.[4]

[3] As will become clear later, the treatments in Cox et al. (2016) are all located in area 11 of Fig. 2 while we expose subjects to decision situations in each of the cells in the figure.

[4] Our design can be seen as a (generalization of a) hybrid between an *investment game* (à la Berg et al. 1995) where both players have rich choice sets (provided the FM has made a "trusting choice") and a *mini trust game* (à la McCabe et al. 2003) where both players have only a binary choice to make (provided the FM has made the 'trusting choice'): In our design the FM has a binary choice to make (it

We are interested in how the SM reacts to the opportunities of gains and losses for both players generated by the FM's choice. To investigate this question we expose subjects to a large number of graphical representations of choice situations. Across choice situations we systematically vary the set of opportunity sets available to the FM in the first stage. By doing so, we can investigate how a wide range of "intentions" revealed by the FM's choice affect the SM's benevolence in the second stage.

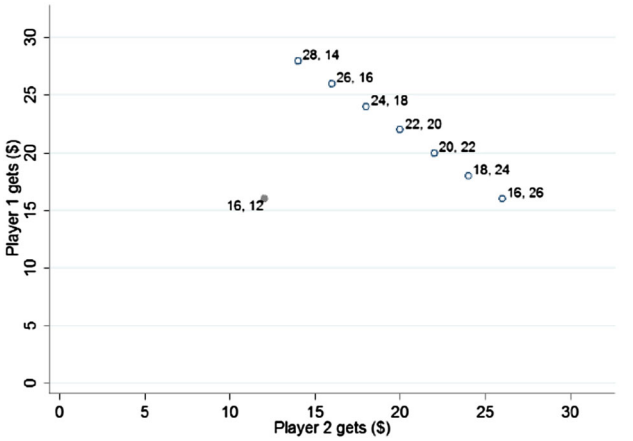
The experiment was conducted by pencil and paper with students from a large Australian university. The subjects in the experiment were recruited via the ORSEE software by Greiner (2015). After subjects read the instructions (they are contained in the online appendix), they were read aloud by an experimenter. Subjects then had to answer a series of control questions to assure their understanding of the task and the payoff procedure. Subjects who answered one or more of the questions incorrectly were informed that the answer was incorrect and the experiment did not proceed before all subjects had answered all control questions correctly. Then each participant was randomly assigned a role, either the role of a FM or the role of a SM. The randomization was such that in each session we had the same number of FMs and SMs and the participants kept their roles during the entire session.

Subjects in both roles were faced with 60 graphical representations of sets of opportunity sets. Each set of opportunity sets consisted of two options, a singleton opportunity set consisting of a fixed payoff-allocation and a non-trivial opportunity set consisting of seven possible payoff-allocations threaded on a downward sloping straight line. In the following we call the former option *the point* and the latter *the line*. If the FM chooses *the point* the SM has no further move while for *the line* she has to decide among the seven allocations in the non-trivial opportunity set. To obtain data from all SMs we used the strategy method: Each SM was asked to make a decision as if the paired FM had chosen the non-trivial opportunity set.[5]

Figure 1 shows a typical example of a decision situation. The task of the FM (Player 1) is to check one of the boxes below the figure indicating whether he prefers option A, *the point*, or option B, *the line* of hollow dots. The task of the SM (Player 2) is to indicate her choice by circling her preferred allocation on *the line* of hollow dots. The 60 decision tasks differed in the positions of the available opportunity sets and the positions were allocated randomly to pairs of subjects. The

Footnote 4 continued
can be interpreted as a choice between transferring a given amount s to the SM and not transferring anything), while the SM has a richer choice set (in our design a choice between seven allocations provided the FM has transferred s). Some of the games investigated by Charness and Rabin (2002) constitute special cases of our design. They found in these cases that the SM often reciprocated to the kindness of the FM (as revealed by his choice). Our design systematically varies the set of choices offered to the FM to investigate other potential factors driving the behavior of the SM.

[5] While there are potential effects of using the strategy method instead of the direct-response method (such as a reduction in incentives or a "hot" versus "cold" effect that might affect the participants' choices—see Zizzo 2010, for a discussion), the experimental literature reports no case in which a treatment effect was observed with the strategy method and not with the direct-response method (see Brandts and Charness 2011).



Player 1: Please tick your preferred option
☐ Option A: The filled dot
☐ Option B: The line of hollow dots (Let Player 2 make a choice)

Player 2: Please choose one of the allocations on the line of hollow dots by circling it in the figure.

Fig. 1 Typical decision task

randomization ensured that *lines* stayed in the positive orthant and the location of *the point* was varied around the *lines* as depicted in Fig. 2.[6,7]

In each session, one subject in the role of the FM and one subject in the role of the SM received exactly identical experimental questionnaires—that is, two experimental subjects in each session faced exactly the same 60 decision tasks. At the end of the experiment we paired the subjects who received identical questionnaires. In each pair we then picked randomly one of the 60 decision tasks, and paid the participants the payoffs corresponding to their joint choices in this situation. Overall, sessions lasted around one hour and participants earned AUD 16.5, on average, plus a show up fee of AUD 5.

3 Conceptual framework

3.1 Second mover's social preferences

In line with the revealed intentions approach, we suppose that the SM cares about how the opportunity set chosen by the FM compares to the alternative opportunity set the FM could have chosen instead. Similar to Cox et al. (2016) we extend this approach by not only looking at the possible gains for both players but also at the possible losses. Compared to Cox et al. (2016) we study a richer array of possible

[6] The randomization also limits concerns for an indirect experimenter demand effect whereby participants observing systematic variations of the location of a *point* relative to the same *line* would infer that their behavior is expected to change as a consequence of the relative position of *the point*.

[7] See online appendix for further details.

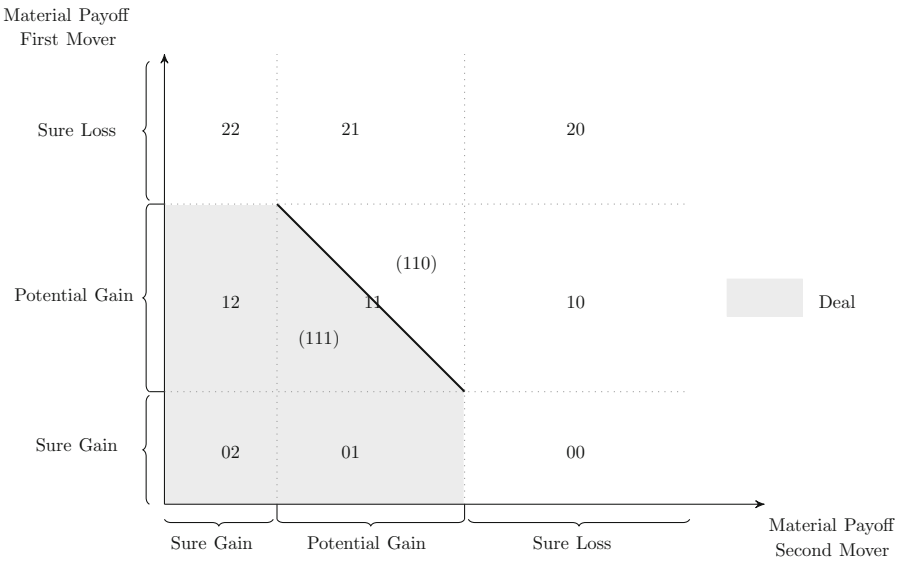


Fig. 2 Observable characteristics of the FM's choice when choosing *the line* for different positions of *the point* relative to *the line*

motivations covering all constellations displayed in Fig. 2.[8] We discuss the features of the areas in this figure in the next subsection.

To allow for errors in decision making, we adopt a random utility approach. In our experiment, in each of the 60 decision tasks, the SM's opportunity set consists of seven discrete options. We therefore use a random-utility discrete choice framework—see Train (2009) for details.

Experimental data from dictator games suggests that the egocentric altruism model by Cox and Sadiraj (2007) or a similar constant-elasticity-of-substitution utility function represents revealed preferences quite well (see Andreoni and Miller 2002 or Cox and Sadiraj 2012, for instance). To incorporate reciprocal motivations, Cox et al. (2007) extend the egocentric altruism model by allowing an agent's willingness to pay for increases or decreases in the payoff of another person (hereafter "benevolence") to depend on this other person's prior actions (that is, on whether the other person was kind or harmful to the agent). Specifically, Cox et al. (2007) propose a model where a subject's benevolence depends on her emotional state, which in turn depends on the other player's choice. For the two-player case the proposed utility function reads:

[8] Cox et al. (2016) design comprises five treatments implemented between subjects. In all these treatments the SM decides how to divide 60 experimental currency units between herself and the FM in case the FM sends her his endowment of 15. The treatments differ in what happens in case the FM decides not to send the endowment to the SM, and whether the FM can make such a decision at all. Thus, in the language of the current paper, the Cox et al. (2016) design keeps the location of the *line* constant and varies the location of the *point* and whether a *point* is available at all. In terms of Fig. 2, the Cox et al. design only investigates constellations in area 11 while we expose subjects to decision situations in each of the cells in the figure.

$$u(x_s, x_o) = \begin{cases} \left(x_s^{\alpha} + \theta x_o^{\alpha}\right)\alpha^{-1} & \alpha \in (-\infty, 0) \cup (0, 1] \\ x_s x_o^{\theta} & \alpha = 0, \end{cases} \tag{1}$$

where x_s is the subject's own material payoff which contributes positively to her utility, x_o is the payoff of the other subject and α and θ are parameters, both supposed to be (weakly) smaller than one. The parameter θ is called the agent's "emotional state" and the effect of the other's payoff on utility depends on the sign of θ. A positive θ means that the individual under consideration cares positively for the other agent in the sense that she is willing to give up money to increase the other's payoff. The agent's willingness to pay—which is the amount of own income the agent is willing to give up in order to increase the other agent's income by one unit—is given by:

$$WTP = \frac{\delta u/\delta x_o}{\delta u/\delta x_s} = \theta\left(\frac{x_s}{x_o}\right)^{1-\alpha}.$$

As is easily seen, the larger θ, the higher the WTP. Note further that α measures the importance of relative payoffs. For positive θ, $\alpha = 1$ yields linear preferences implying that the WTP is independent of relative payoffs, while $\alpha < 1$ yields convex preferences implying that the WTP and with it the agent's benevolence towards the other agent increases as the other's relative payoff decreases.

Here we adopt this functional form and—in line with Cox et al. (2007)—we capture intention-based benevolence from the SM by allowing her emotional state θ to depend on the FM's previous choice. Specifically, we allow a SM's θ to depend on the observable characteristics of the choice of the FM as defined in the next subsection:

$$\theta = \theta(\text{observable characteristics of the FM's choice}).$$

3.2 Classification of first mover's choices, attributed intentions and their impact on second movers' behavior

The classification in Fig. 2 is based on the gain/loss principle applied to both players, i.e. whether the opportunity set that was chosen by the FM (*the line*) comes with an actual or potential increase or decrease of each player's payoff compared to the not chosen opportunity set (*the point*). Our first hypothesis is motivated by the experimental evidence indicating that reciprocity is an important driver for behavior in games. Reciprocation entails responding to positive perceived kindness with positive kindness, and to negative perceived kindness with negative kindness (Rabin 1993; Charness and Rabin 2002; Dufwenberg and Kirchsteiger 2004). In a material context kindness is usually equated with generosity. To formulate a hypothesis regarding the impact of positive reciprocity on the behavior of the SM we therefore characterize the choice of the FM in terms of the implied generosity towards the SM. Here we distinguish between three levels of *generosity* when the FM chooses *the line* over *the point*:

Definition 1 Let the FM choose the actual opportunity set for the SM from a collection consisting of a point and a line. Suppose the FM chooses the line.

(a) If the chosen opportunity set (*the line*) only includes allocations which decrease the SM's payoff compared to the not chosen opportunity set (*the point*), the FM's choice is said to imply a **sure loss for the SM**.
(b) If *the line* includes allocations for which the SM's payoff is (weakly) higher, and allocations for which the SM's payoff is (weakly) lower compared to her payoff in *the point*, the FM's choice of *the line* is said to imply a **potential gain for the SM**.
(c) If *the line* only includes allocations which (weakly) increase the SM's payoff compared to *the point*, the FM's choice of *the line* is said to imply a **sure gain for the SM**.

Using this classification of FM behavior, it seems plausible that choices of the FM that imply a sure gain for the SM are interpreted by the SM as more generous than choices that imply a potential gain for the SM, and that choices that imply a potential gain for the SM are interpreted as more generous than choices that imply a sure loss for the SM. This consideration yields our first prediction:

Hypothesis 1 (*Impact of generosity*) The SM's benevolence increases with the level of generosity implied by the choice of the FM. That is, the SM becomes progressively more benevolent when we move from situations where the FM's choice implies a sure loss for the SM, to situations where the FM's choice implies a potential gain for the SM, to situations where the FM's choice implies a sure gain for the SM.

Our second hypothesis is based on experimental evidence indicating that the vulnerability of the FM is an important driver for the behavior of the SM in the investment game (see Cox et al. 2016 for an investigation of the role of vulnerability in the investment game). Vulnerability in our context means that the FM, by choosing the *line*, accepts the risk of losing money depending on the SM's choice. To formulate a hypothesis regarding the impact of vulnerability on the behavior of the SM we therefore characterize the choice of the FM in terms of the implied risk for the FM. Here we distinguish between three levels of *vulnerability* of the FM when he chooses *the line* over *the point*:

Definition 2 Let the FM choose the actual opportunity set for the SM from a collection consisting of a point and a line. Suppose the FM chooses the line.

(a) If the chosen opportunity set (*the line*) assures the FM a payoff increase compared to the not chosen opportunity set (*the point*), the FM's choice is said to imply a **sure gain for the FM**.
(b) If *the line* includes allocations for which the FM's payoff is (weakly) higher, and allocations for which the FM's payoff is (weakly) lower compared to the payoff in *the point*, the FM's choice of *the line* is said to imply a **potential gain for the FM**. In that case we also say that the FM's choice of *the line* makes him **vulnerable**.

(c) If *the line* only includes allocations which decrease the FM's payoff compared to *the point*, the FM's choice of *the line* is said to imply a **sure loss for the FM**. In this case we also say that the FM's choice of *the line* corresponds to a **sacrifice**.

Using this classification of FM behavior we now posit two hypotheses. Hypothesis 2a predicts that choices by the FM that make him vulnerable lead to benevolent behavior by the SM:

Hypothesis 2a (*Impact of vulnerability*) The SM's benevolence increases if the FM's choice makes him vulnerable. Specifically, the SM becomes more benevolent when we move from situations where the FM's choice implies a sure gain for the FM, to situations where the FM's choice implies a potential gain for the FM.

We also suspect that FM choices that correspond to a sacrifice influence the behavior of the SM. This is the content of Hypothesis 2b. Note that Hypothesis 2b does not make any prediction on how the effect of sacrifice compares to the effect of vulnerability.

Hypothesis 2b (*Impact of sacrifice*) The SM's benevolence increases if the FM's choice implies a sacrifice for him. Specifically, the SM becomes more benevolent when we move from situations where the FM's choice implies a sure gain for the FM, to situations where the FM's choice implies a sure loss for the FM.

Our next hypothesis is based on the idea that SMs may reward FM choices that have the potential to increase the payoffs of both parties. This conjecture is motivated by the experimental evidence indicating that efficiency concerns are important for behavior in the lab and in the field (see Engelmann and Strobel 2004; Fehr et al. 2006, among others). To formulate a hypothesis regarding the impact of efficiency concerns on SM behavior we characterize FM choices according to the payoff consequences for both players as follows:

Definition 3 Let the FM choose the actual opportunity set for the SM from a collection consisting of a point and a line. Suppose the FM chooses *the line*. If *the line* includes allocations which represent a Pareto improvement relative to *the point*, the FM's choice is said to allow for a **deal**.

We then state:

Hypothesis 3 (*Impact of deal*) The SM's benevolence increases if the FM's choice allows for a deal. That is, the SM becomes more benevolent when we move from situations where the choice of the FM does not allow for a Pareto improvement to situations that allow for a mutual improvement.

Our next (and last) hypothesis is motivated by the large experimental literature on trust and trustworthiness. In experimental economics the most frequently used instrument to study the importance of those concepts for behavior is the investment game (Berg et al. 1995) and its close relative, the binary trust game (studied by McCabe et al. 2003, for instance). There is by now an impressive amount of

evidence indicating that SM behavior in those games is neither consistent with own money maximization nor in line with purely distributional concerns (see Cox 2004; Ashraf et al. 2006; Chaudhuri and Gangadharan 2007; Cox et al. 2008a, 2016, among others). Less clear is the answer to the question what is really driving SM behavior in this class of games. Here, we address this question indirectly by investigating whether FM behavior characterized by the combination of characteristics defining a trusting move in the investment game induces more benevolence in the SM than behavior characterized by other combinations. To formulate a hypothesis regarding the impact of trusting acts by the FM on the behavior of the SM we define:

Definition 4 Let the FM choose the actual opportunity set for the SM from a collection consisting of a point and a line. Suppose the FM chooses *the line*. If the choice of *the line* makes the FM vulnerable and if in addition it allows for a deal, then the FM's choice is said to reveal **trust**.

We then hypothesize that choices revealing trust have the power to trigger benevolence in the SM:

Hypothesis 4 (*Impact of trust*) The SM's benevolence increases if the FM's choice reveals trust. That is, the SM becomes more benevolent when we move from situations where the choice of the FM does not reveal trust to situations where the choice of the FM reveals trust.

4 Data and results

We first provide an overview of the data collected in our experiment and a descriptive analysis. We then proceed with the parameter estimation of our model.

4.1 Data

We carried out 14 experimental sessions involving 190 subjects in total. Since our research focus lies on the conditional part of an individual's social preferences, we are only interested in the data collected from experimental SMs. Since we collected the data via the strategy method, our data set consists of 60 decisions for each of the 95 SMs.

Looking at the individual data, we find that 37 subjects (that is, 38.9 percent of our SM population) behaved in a perfectly selfish way by choosing the right-most point on *the line* in each of the 60 decision situations. Hence, $\theta = 0$ and $u(x_s, x_o) = x_s$ for almost 40 percent of our SM sample. This is in line with empirical evidence presented in related research—by Fehr and Gächter (2000), and Andreoni and Miller (2002), for instance), where between 20 and 50 percent of the individuals are found to act in a completely selfish manner.

For our further analyses, we exclude the purely selfishly acting SMs from our data sample and focus on the 58 participants that reveal some form of other-

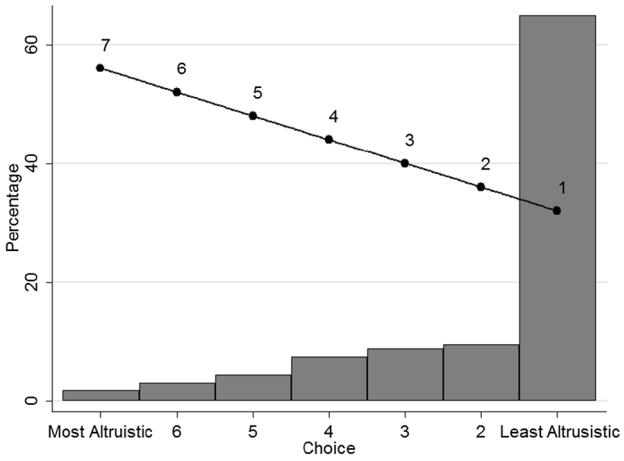


Fig. 3 SM's choice distribution

Table 1 Summary of participants' choices

Choice	Freq.	%	Cum. (%)
7 (Most altruistic)	106	3.1	3.1
6	177	5.1	8.2
5	252	7.3	15.4
4	422	12.2	27.6
3	498	14.4	42.0
2	540	15.6	57.6
1 (Least altruistic)	1468	42.4	100.0
Total	3463	100.0	

regarding behavior.[9] The overall distribution of the choices of those SMs is presented in Fig. 3 and Table 1.[10] In Table 1 we see that the left-most four points on *the line* (points 4–7) are chosen in only 27.6 percent of the decision tasks. This is not really surprising, as point 7 is the most benevolent decision a SM can make, and point 1 is the least benevolent one. Thus, the subjects in the subsample under consideration—although not purely selfish—have a tendency to care more for their own than for the other's payoff.

4.2 Descriptive analysis

In a first step, we analyze whether the characteristics defined in Sect. 3 influence SM behavior. For this purpose we first define a set of binary variables reflecting

[9] Since $\theta = 0$ for purely selfishly acting individuals, the behavior of subjects in this subsample is not informative about how intentions influence social preferences.

[10] The experiment was conducted by pen and paper and a small number of answers ($N = 17$) were missing in the questionnaires. This leaves a dataset of 3463 observations.

Definitions 1 and 2 introduced in Sect. 4: "sure gain for the FM" (FM_{SG}), "potential gain for the FM" (FM_{PG}) and "sure loss for the FM" (FM_{SL}), as well as "sure gain for the SM" (SM_{SG}), "potential gain for the SM" (SM_{PG}) and "sure loss for the SM" (SM_{SL}). In addition to the effects of these binary variables, we analyze the effect of the dummy "*Deal*", which is one if the choice of *the line* allows for a deal according to Definition 3 and zero otherwise; and we also analyze the effect of the dummy "*Trust*", which is one if the choice of *the line* reveals trust according to Definition 4 and zero otherwise.[11]

Our first observation supports our main hypothesis that the choice of the SM on *the line* depends significantly on the nature of the counterfactual choice the FM could have made: Fig. 4 displays the mean SM choice as a function of the characteristics of the FM's choice. The significance of the difference in means is indicated using *t*-tests (from regressions on dummies using cluster robust variance to control for the non-independence of observed choices within participants). The bars in the upper-left panel of Fig. 4 suggest that the choices of SMs become more benevolent if the level of generosity increases from SM_{SL} to SM_{PG} ($p = 0.072$) and from SM_{PG} to SM_{SG} ($p = 0.010$). The bars in the upper-right panel of Fig. 4 suggest that SMs also become more benevolent if the FM's choice implies vulnerability—moving from FM_{SG} to FM_{PG} ($p = 0.001$)—or sacrifice—moving from FM_{SG} to FM_{SL} ($p = 0.012$). Interestingly, the mean choice of SMs is not significantly different between situations characterized by FM_{PG} and situations characterized by FM_{SL} ($p = 0.437$). Turning to *Deal* and *Trust* in the lower part of Fig. 4 we find that SMs are relatively more benevolent when the choice of the FM allows for a *Deal* ($p = 0.036$) or reveals *Trust* ($p = 0.027$). It should be noted, however, that this latter observation does not imply that SMs react to *Deal* and *Trust* per se; they might rather react to the FM's generosity and vulnerability which are both present in situations of *Deal* and *Trust*.

The effect of the counterfactual choice the FM could have made on SM's behavior can also be seen in Fig. 5. In this figure the cumulative distribution functions (CDFs) of SM choices on the line are represented depending on the level of generosity, the level of vulnerability, and on whether the choice of the FM allows for a *Deal* or reveals *Trust*. A first-order stochastically dominating CDF reflects more benevolence. It can be seen that the CDF for choices exhibiting SM_{SG} first-order stochastically dominates the CDF for choices featuring SM_{PG} (KS test: $p = 0.025$), which in turn first-order stochastically dominates the CDF for FM choices featuring SM_{SL} (KS test: $p = 0.005$). This finding strengthens the previous result that a more generous choice by the FM triggers a more benevolent response by the SM, and therewith provides further support for Hypothesis 1.

We also find support for Hypothesis 2. The CDF of choices featuring FM_{PG} first-order stochastically dominates the CDF of choices with FM_{SG} (KS test: $p = 0.003$), which is clearly in line with Hypothesis 2a. It is also the case that the CDF of choices featuring FM_{SL} first-order stochastically dominates the CDF of choices with FM_{SG} (KS test: $p = 0.034$), which is in line with Hypothesis 2b. Comparing the

[11] Note that the shaded areas in Fig. 2 cover situations where the choice of *the line* allows for a deal, while area 111 contains all situations where the choice of *the line* reveals trust.

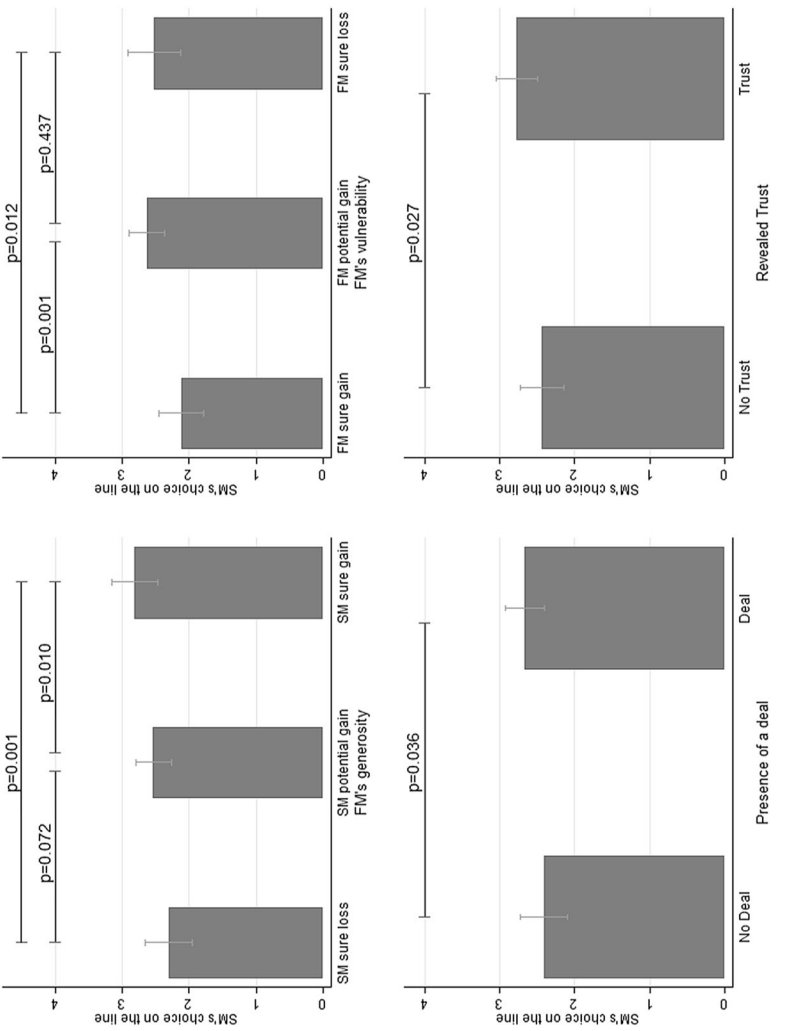


Fig. 4 SM's benevolence as a function of the characteristics of the FM's choice. The *figure* shows the average choice of the SM on the *line*. *Higher values* indicate more benevolence

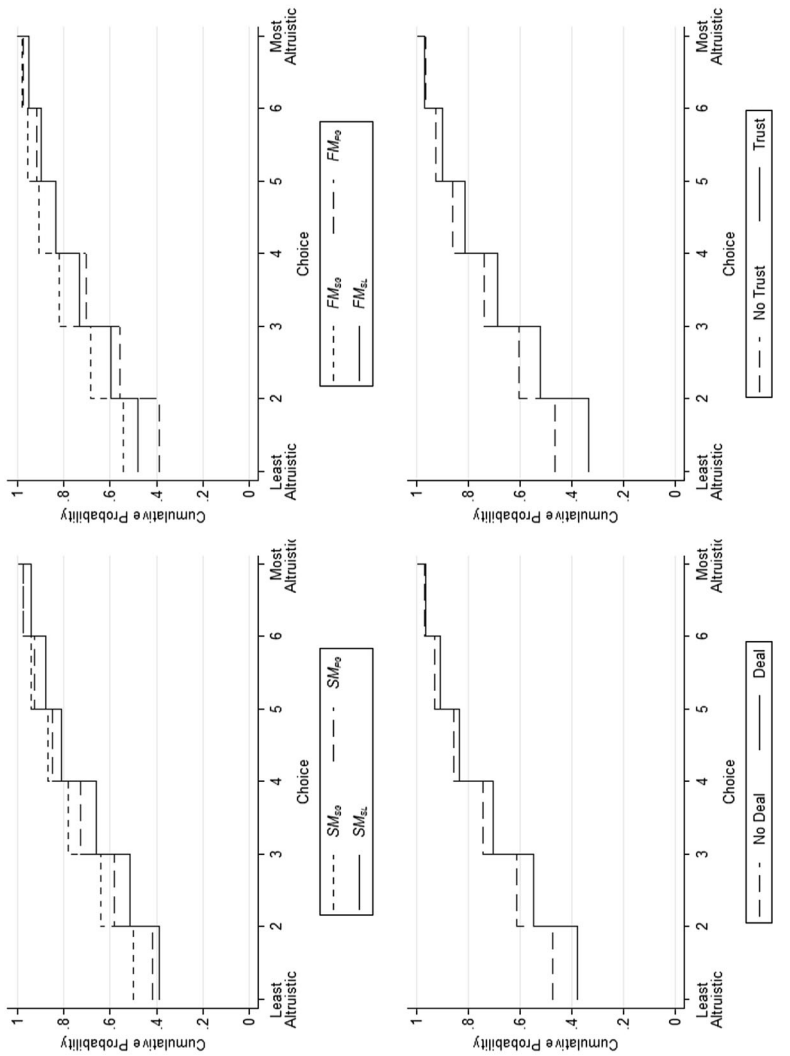


Fig. 5 Cumulative distributions of the SM's choice by characteristics of the FM's choice

distribution of choices featuring FM_{PG} to the distribution of choices featuring FM_{SL} we see that they differ (KS test: $p = 0.001$) although the mean choice is statistically indistinguishable between the two situations. Specifically, the distribution of SMs' responses to FM_{SL} features both more most altruistic choices and more least altruistic choices. In the lower part of Fig. 5 we see that the CDF of choices that allow for a *Deal* first-order stochastically dominates the CDF of choices without a *Deal* available (KS test: $p < 0.001$). However, as previously stated this finding might be confounded by the fact that if the FM's choice allows for a *Deal*, it necessarily also entails either SM_{SG} or SM_{PG}—which might be responsible for the effect on SM's benevolence. Finally, we also find some support for Hypothesis 4: The CDF of SM choices featuring *Trust* almost first-order stochastically dominates the CDF of SM choices not revealing *Trust* (KS test: $p < 0.001$). Here again, this finding might be confounded by the fact that the SM may simply react to the generosity and vulnerability which characterize the trust situation.

4.3 Disentangling revealed intentions

The structural model described in Sect. 3.1 makes it possible to disentangle the effects of different characteristics of the FM's choice on the SM's behavior. Following the random utility approach (Train 2009), we assume that the utility of SM i for payoff pair $x = (x_s, x_o)$ in a choice situation featuring the characteristic combination j includes a stochastic term which represents the unobserved part of utility (including quixotic variations in utility due to cognitive limitations when assessing the options):

$$v_j^i(x) = \left(x_{SM}^{\alpha} + \theta_j^i x_{FM}^{\alpha}\right)\alpha^{-1} + \varepsilon, \tag{2}$$

with

$$\begin{aligned} \theta_j^i = \theta_{00} + \theta^i \\ + \beta_{FM_{PG}} \mathbb{1}_{j\in\{10,110,111,12\}} + \beta_{FM_{SL}} \mathbb{1}_{j\in\{20,21,22\}} \\ + \beta_{SM_{PG}} \mathbb{1}_{j\in\{01,110,111,21\}} + \beta_{SM_{SG}} \mathbb{1}_{j\in\{02,12,22\}} \\ + \beta_D \mathbb{1}_{j\in\{01,02,111,12\}} + \beta_T \mathbb{1}_{j=111} \end{aligned} \tag{3}$$

Here, θ_j^i is the emotional state of SM i when she observes that the FM has chosen *the line* in a choice situation where the alternative choice he could have made (that is, *the point*) is located in area $j \in \{01, 02, 10, 110, 111, 12, 20, 21, 22\}$ as defined in Fig. 2. This formulation assumes that "motives are additive" in the sense that adding a given motive has the same effect independently of whether other motives are present or absent. We will relax this assumption later on. Note also that this formulation allows for individual heterogeneity in social preferences with the inclusion of an individual specific term θ^i. We follow a standard approach in discrete choice modeling (Train 2009) in assuming $\varepsilon \rightsquigarrow Gumbel(\lambda)$. This implies that the choice model is a non-linear multinomial logit model with the probability that a given allocation x' is chosen among a set X of possible allocations given by:

Table 2 Estimation of α and θ_j^i by maximum-likelihood taking FM_{SG} and SM_{SL} as reference categories

Model ($N = 3463$)		$v_j^i(x) = \left(x_{SM}^{\alpha} + \theta_j^i x_{FM}^{\alpha}\right)\alpha^{-1} + \varepsilon$	
Parameter		Estimate	Robust SE
α		0.282	0.170
θ	FM payoffs		
	FM_{SL}	0.188*	0.080
	FM_{PG}	0.190**	0.071
	FM_{SG}	(Ref)	
	SM payoffs		
	SM_{SG}	0.206*	0.098
	SM_{PG}	0.042	0.038
	SM_{SL}	(Ref)	
	Deal	0.001	0.066
	Trust	0.027	0.075
λ		4.078**	1.391

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

$$\mathbb{P}(x') = \frac{\exp(\lambda v(x'))}{\sum_{x \in X} \exp(\lambda v(x))},$$

where λ is the subjects' precision parameter.[12] We estimate the parameters α and θ_j^i by maximum-likelihood. Each participant provided 60 data points, we therefore cluster the standard error by participant.

Table 2 reports the estimates of our basic model. As expected $\alpha < 1$ which indicates convex preferences. In the sequel we focus our discussion on the parameter θ_j since this is the parameter related to our research question. The impact of the characteristics of the FM's choice on this parameter is measured in comparison to the reference categories FM_{SG} and SM_{SL}, respectively. These reference categories are arguably associated with the lowest level of benevolence by the SM.

The parameter estimates in Table 2 suggest that—starting from the reference categories—an increase in the level of generosity from the FM towards the SM, as well as an increase in the FM's vulnerability indeed have a significant positive impact on the SM's altruism coefficient θ and thus on her benevolence. Regarding generosity, we find that a sure gain for the SM (SM_{SG}) has a significant effect on the SM's benevolence while a sheer potential gain (SM_{PG}) does not have a significant effect. This result provides partial support for Hypothesis 1:

Result 1 (Impact of generosity) *The SM's altruism coefficient θ and therewith her benevolence increases with the level of generosity implied by the choice of the FM. However, the effect is significant only for situations where the FM's choice implies a sure gain for the SM.*

Turning to the effect of vulnerability, we see that FM_{PG} raises θ significantly. This result confirms the finding of the descriptive analysis and is in line with

[12] This is called the "Luce model" (see Wilcox 2008).

Hypothesis 2a. In line with Hypothesis 2b we also find a positive effect of FM_{SL} on the SM's benevolence. Comparing the two we see that the estimated coefficients of FM_{SL} and FM_{PG} are roughly equal. Thus, acts that make the FM vulnerable and acts that imply a sure loss for the FM seem to have a similar impact on the intention-perception of the SM as revealed by her behavior.

Result 2 (Impact of vulnerability and sacrifice) *The SM's altruism coefficient θ and therewith her benevolence increases if the choice of the FM entails vulnerability (potential loss for the FM) or sacrifice (sure loss for the FM). Comparing the two effects we find that they are similar in size.*

Turning to the question of whether the behavior of SMs becomes more benevolent when the choice by the FM allows for a Pareto improvement, we observe that *Deal* availability has no significant effect on the benevolence of the SM. Hypothesis 3 is therefore not supported by the data. The previously observed shift in the CDF of SM's choices (Fig. 5) seems indeed to be driven by generosity or vulnerability.

Result 3 (Impact of deal) *The availability of a deal by itself has no effect on the SM's altruism coefficient θ and therewith on her benevolence.*

Similarly, we do not observe any effect of trust in itself when the potential gains and losses of the two players are controlled for. Hypothesis 4 is therefore not supported by the data either. Here again the shift in the CDF of the SM's choices between situations where the FM's choice reveals trust and situations where it does not (Fig. 5) seems to be driven by the effects of generosity and vulnerability without an additional impact of trust in itself.

Result 4 (Impact of trust) *The expression of trust has no effect in itself on the SM's altruism coefficient θ and therewith on her benevolence.*

As previously mentioned our estimation of model (2) assumes that motives are additive—see Eq. (3). We now relax the additivity assumption and allow for possible interactions between the FM's vulnerability and his generosity towards the SM. Specifically, we define a dummy for each area displayed in Fig. 2 and estimate the model:

$$\theta_j^i = \theta_{00} + \sum_k \beta_k \mathbb{1}_{j=k} + \theta^i \tag{4}$$

with $j, k \in \{01, 02, 10, 110, 111, 12, 20, 21, 22\}$.

Our chosen reference category is again $FM_{SG} \times SM_{SL}$ (area 00 in Fig. 2). The estimation results of this model are presented in Fig. 6. By and large the results confirm our earlier findings. We observe that the SM's benevolence is high in situations where the FM's choice makes him vulnerable as long as vulnerability comes together with either a potential or a sure gain for the SM (SM_{SG}, SM_{PG}). The SM's benevolence is also always significantly positive for situations where the FM's choice implies a sacrifice, and, as long as the choice implies either a potential or a sure gain for the SM there is no significant difference between the reaction of the

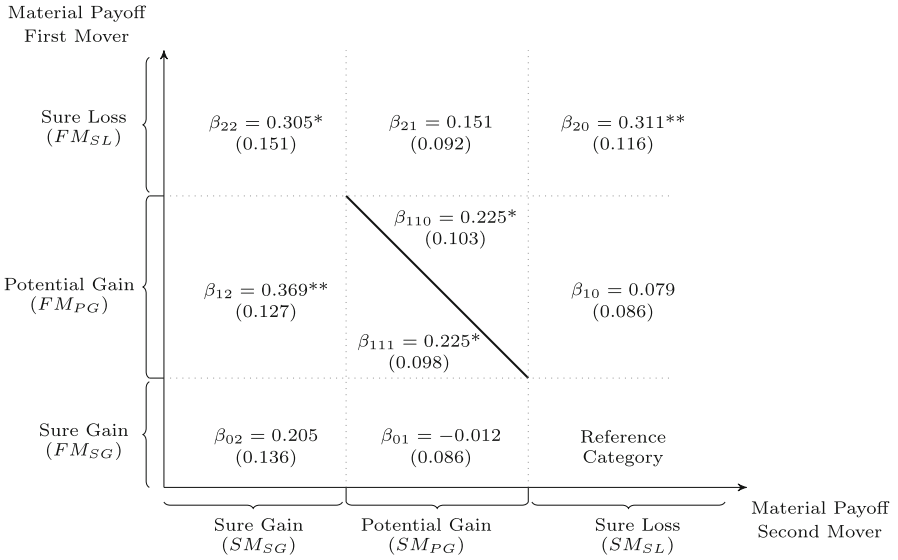


Fig. 6 Maximum-likelihood estimation results of θ_j. The chosen reference category is $FM_{SG} \times SM_{SL}$. Robust standard errors are displayed in *brackets*. Significance: $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$

SM to vulnerability and her reaction to sacrifice (no significant differences between β_{12} and β_{22} and between β_{110}, β_{111} and β_{21}). When the FM's choice implies a sure loss for the SM (SM_{SL}), the SM's benevolence increases with the opportunities of losses for the FM: FM_{PG} has a positive but insignificant effect and FM_{SL} has a positive and significant effect. This latter effect seems rather strange at first sight and it is investigated further in the next subsection.

While SM_{PG} and FM_{PG} in isolation are not enough to influence the benevolence of the SM (β_{01} and β_{10} are not significantly different from zero), it is noteworthy that their joint presence (in β_{110}, and in β_{111}) is. It therefore looks like there is an interaction between the effect of generosity and the effect of vulnerability. Increasing the level of generosity to SM_{SG} enhances the SM's benevolence even further (β_{12} is significantly larger than β_{111} and β_{110}), assuring the highest level of benevolence by the SM observed in our experiment.

Turning to Hypotheses 3 and 4 about the impact of *Deal* and *Trust* on the behavior of the SM we see that the coefficients β_{110} and β_{111} do not significantly differ from each other. Area 111 corresponds to FM choices revealing *Trust* and it differs from area 110 by the presence of a *Deal*. Thus, the relatively high level of benevolence from the SM observed in the area 111 seems solely be driven by the presence of FM_{PG} and SM_{PG}. This finding supports and strengthens our previously stated Results 3 and 4.

Is model (4) necessary? As we estimate many parameters, it may be that some interactions happen to be positive by chance. Using a Fisher test of joint significance we can assess whether the extra parameters in (4) relative to (3) are significant. We find that indeed we cannot accept the null hypothesis according to which the coefficients in model (4) can just be generated by model (3) where there are no

interactions ($p = 0.013$). Our results therefore suggest that the outcomes for the FM and the SM interact in influencing the SM's view about the FM's intentions.

That being said, overall we conclude that relaxing the hypothesis that motives are additive does not change our previous results qualitatively: Positive reciprocity—whereby a generous choice by the FM triggers a benevolent response by the SM—and vulnerability-responsiveness—whereby a choice by the FM that exposes him to the risk of losing money triggers a benevolent response–seem to be important drivers for SM behavior, while deal-responsiveness—where the SM reacts positively to choices that create the possibility of mutual improvements—or trust-responsiveness—where the SM rewards acts that reveal trust—seem behaviorally less relevant.

4.4 Interpreting intentions from observed actions and salient social norms

In the precedent analyses, we have investigated whether a SM's benevolence is affected by the objective characteristics of the FM's choice—specifically by how his actual choice compares to the counterfactual alternative choice he could have made instead. By doing so we have extended the revealed intention approach and looked at the possible gains and losses created by the FM's decision. Here we argue that this approach can be extended further by incorporating the possible role of preexisting *social norms* in the analysis. Social norms are by definition shared and common knowledge (Krupka and Weber 2013). In games where allocations of resources are made between players, prevailing social norms may point to a "fair" allocation, that is, one which would be considered as such by the different players. In an experiment where subjects enter the laboratory as equals, where they are allocated randomly to their roles and where the money to be divided is a windfall provided by the experimenter, it seems plausible that fairness norms point to an equal split. Even though equal sharing might not be the only norm prevalent in the population of experimental subjects (e.g., asymmetry of roles may be considered as giving different entitlements to different players), it is likely to be the most prevalent norm.

A look at the choices of experimental SMs suggests that the equality norm has indeed an impact. Figure 7 shows by how much the SM's choice differs from the least unequal allocation (the feasible allocation on *the line* that is closest to the 45 degree line, henceforth LUA).[13] Positive (negative) entries in Fig. 7 correspond to choices on *the line* where the SM earns more (less) in material terms than the associated FM. As can be seen from the figure there is a large concentration of SM choices at the LUA (more than a third of all choices by experimental SMs are at the LUA) and there is a pronounced discontinuity in the distribution of choices immediately to the left of the LUA, arguably because there is no social norm that dictates to give more than the "fair share" (implied by the LUA) to the other player. It therefore seems that the 50–50 split indeed plays a role for SM behavior.

[13] If *the line* crosses the 45 degree line and if the crossing point is one of the seven feasible allocations on *the line*, then this allocation is the LUA; if *the line* crosses the 45 degree line but the crossing point is not a feasible allocation then the feasible allocation on *the line* that is closest to the 45 degree line is the LUA; and if *the line* does not cross the 45 degree line then the feasible allocation on *the line* that is closest to the 45 degree line is the LUA.

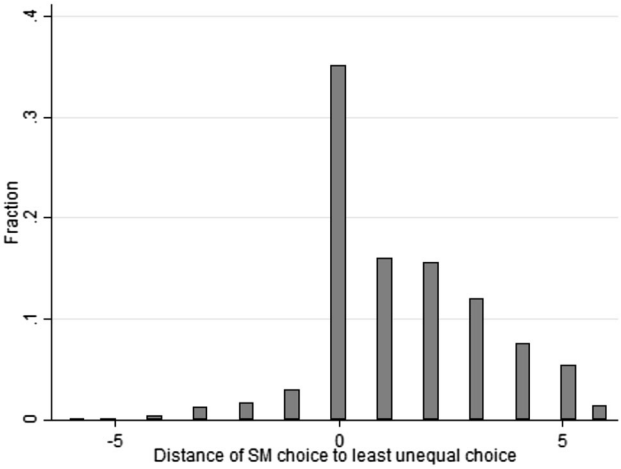


Fig. 7 Distribution of the distance between the actual choice of the SM and the least unequal allocation on the line. *Positive numbers* represent unequal choices in favor of the SM, *negative numbers* represent unequal choices in favor of the FM

We next ask whether the interpretation of the FM's intentions by the SM is influenced by this norm. To address this question we extend the revealed intentions approach by investigating whether SM choices are affected by the fairness of the counterfactual choice, *the point*, taking equality as the yardstick. Specifically, we estimate θ_j^i, using Eq. (2), separately for situations where *the point* is above the 45 degree line and situations where it is below that line. Table 3 displays the associated parameters. It shows that in both sub-samples coefficients have the same sign as reported for the aggregate data, but the parameters are smaller and not significant when *the point* is an allocation that favors the FM, while the coefficients of FM_{SL} and SM_{SL} are relatively large and significant when *the point* is to the advantage of the SM. Overall this result suggests, that intentions are read in relation to the 50–50 social norm. The SM reacts more positively to the generosity of the FM and to his sacrifice, when the FM chooses *the line* in a situation where *the point* is an allocation characterized by inequality in favor of the SM. One interpretation of the result that the SM reacts more benevolently to the generosity of the FM in a situation where *the point* is an allocation characterized by inequality in favour of the SM is that in such a situation—by choosing *the line*—the FM is offering gains to the SM even though the SM was already advantaged by the initial allocation. Similarly, in a situation where *the point* is an allocation characterized by inequality in favour of the SM and where the choice of *the line creates* a sure loss for the FM, the choice to sacrifice might be considered as particularly noticeable by the SM because the FM was already disadvantaged by the initial allocation.

Turning to the result that the SM is relatively benevolent in the $FM_{SL} \times SM_{SL}$ situation, we observe that the choice of *the line* by the FM in this constellation can potentially be interpreted as an attempt to avoid a split that is unfavorable to him, even if this leads to a loss in the payoffs of both players. To test whether this

Table 3 Estimation of α and θ_j^i by maximum-likelihood taking FM_{SG} and SM_{SL} as reference categories

Model		$v_j^i(x) = \left(x_{SM}^{\alpha} + \theta_j^i x_{FM}^{\alpha}\right)\alpha^{-1} + \varepsilon$			
Parameter		Start favors FM		Start favors SM	
		Estimate	Robust SE	Estimate	Robust SE
α		0.681*	0.293	−0.135	0.199
θ	FM payoffs				
	FM_{SL}	0.032	0.089	0.181*	0.092
	FM_{PG}	0.070	0.110	0.134	0.082
	FM_{SG}	(Ref)			
	SM payoffs				
	SM_{SG}	0.074	0.100	0.275*	0.118
	SM_{PG}	0.021	0.045	0.065	.049
	SM_{SL}	(Ref)			
	Deal	−0.004	0.038	−0.054	0.092
	Trust	0.034	0.052	−0.058	0.086
λ		3.350**	0.893	12.245*	5.934
N		1832		1631	

* $p<0.05$, ** $p<0.01$, *** $p<0.001$

interpretation is consistent with the data, we re-estimate the model (2) allowing for different values of the parameter θ in $FM_{SL} \times SM_{SL}$ situations above and below the diagonal. Table 4 displays the results. Column (1) allows θ to depend on a dummy $Point_{FM}$ taking a value 1 if *the point* favors the FM and zero if it favors the SM (we do not consider situations of equality). We find that the benevolence is overall larger for situations where the FM abandoned a relatively advantageous point when choosing *the line* ($p<0.001$). In column (2), we interact this dummy with a dummy for the $FM_{SL} \times SM_{SL}$ situation. We find that SMs are significantly more benevolent when the $FM_{SL} \times SM_{SL}$ situation appears for *points* below the diagonal. This effect vanishes when the fixed allocation is above the diagonal.

These results are important for the revealed intentions approach. They show that the reaction of the SM to the FM's choice is not only shaped by differences between the opportunities generated by the choice set selected by the FM and the opportunities which could have been generated by a counterfactual choice. The SM's reaction seems also to depend on how a prevailing social norm of equality labels each of these opportunities as fair or not. In the case of our experiment, the puzzling behavior of the SM in $FM_{SL} \times SM_{SL}$ situations makes sense if the SM interprets the FM's choice as an attempt to avoid a split that is unfavorable to him. This, as a consequence, may induce the SM to make a more benevolent choice than in $FM_{SG} \times SM_{SL}$ situations. By contrast, benevolence by the SM is not observed when the (not chosen) point was favorable to the FM.

Table 4 Benevolence as a function of the position of the (not chosen) point relative to the equal-material payoff line

Model ($N = 3203$):	$v_j^i(x) = \left(x_{SM}^{\alpha} + \theta_j^i x_{FM}^{\alpha}\right)\alpha^{-1} + \varepsilon$			
Parameter	(1)		(2)	
	Estimate	Robust SE	Estimate	Robust SE
α	0.288	0.178	0.308	0.179
θ				
$Point_{FM}$	0.170***	0.043	0.177***	0.044
$\mathbb{1}_{FM_{SL} \times SM_{SL}}$			0.254***	0.089
$Point_{FM} \times \mathbb{1}_{FM_{SL} \times SM_{SL}}$			−0.322*	0.153
λ	4.377**	1.497	4.241**	1.435

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5 Discussion

The empirical study of conditional other-regarding preferences based on higher-order beliefs is difficult because such beliefs are not observable and because eliciting them is a tricky task. An elegant alternative to belief-based conditional other-regarding preferences is the revealed intentions approach where a player cares about the generosity of the opportunity set chosen by another player compared to other opportunity sets that could have been chosen. The present paper has extended the revealed intentions approach by allowing agents to care not only about the possibility of gains generated by other agents' actions but also about the possibility of losses. In a two-player two-stage game, we have investigated how the second mover's other-regarding preferences are affected by different characteristics of the opportunity set chosen by the first mover compared to a counterfactual opportunity set the first mover could have chosen. By systematically varying the set of opportunity sets the first mover can choose from and investigating the response of the second mover to the actual choice of the first mover and the alternative choice he could have made, we were able to elicit how the second mover reacts to a wide variety of intentions as revealed by the first mover's choice.

We found that second movers do react to the possibilities of gains and losses generated for them and for the associated first mover. Second movers are typically more benevolent when the choice of the first mover creates an opportunity of gains for the second mover. This can be interpreted as a manifestation of positive reciprocity from the second mover. We have also seen that second movers tend to become more benevolent when the first mover chooses an opportunity set that implies either a potential or a sure loss for him. We interpret this as evidence in support of the hypothesis that vulnerability-responsiveness is an important motivation for second movers. On the other hand, efficiency concerns and an aversion against violating trust seem to be far less important motivations. Finally we found that second movers compare the actual choice of the first mover and the alternative choices that would have been available to him to allocations that involve equal material payoffs.

Interestingly, as pointed to us by a reviewer, our result that second movers tend to be more benevolent in environments where the first mover's choice makes him vulnerable is in sharp contrast to findings by List (2007) and Bardsley (2008) for giving in the dictator game. Specifically, these authors find that fewer dictators are willing to transfer money to the paired recipient when their opportunity set includes actions that correspond to taking from the recipient. A key difference of our work to theirs is obviously that in our experiments the choice of the first mover determines the opportunity set for the second mover while in their dictator games the opportunity set for the dictator is exogenously imposed by the experimenter. Thus, while in our experiments the (endogenous) choice of the opportunity set for the second mover potentially reveals information about the intentions of the first mover, the (exogenous) imposition of the opportunity set of the dictator by the experimenter in their experiments does not reveal such information. Since information about the intentions of one agent might change the preferences of the other player this crucial difference in design is likely to be responsible for the differences in results.

Our result that second movers tend to be more benevolent in situations where the choice of the first mover makes him vulnerable is related to results in recent research investigating the consequences of exerting control in the context of incomplete contracts. In a principal-agent game, by exerting more control over the agent's actions, the principal implicitly reduces the agent's opportunity set—by taking out outcomes that are less profitable for the principal. As Falk and Kosfeld (2006) show experimentally, control entails hidden costs since most agents reduce their performance as a response to the principal's controlling decision. Since electing not to exert control places the principal in a position of vulnerability, this result is similar in vein to our vulnerability-responsiveness result.[14] Future research should certainly investigate further how self-imposed vulnerability affects the intention-based preferences of another player.

Overall, our study shows that it is possible to study a rich array of revealed intentions, without eliciting beliefs, by systematically varying the set of opportunity sets available to the first mover in a two-player two-stage game and by investigating the response of the second mover to the actual choice of the first mover and the alternative choice he could have made instead. Another significant contribution of our study is to show that incorporating a salient social norm, here the equality norm, can be useful to discriminate between different "revealed intentions". Overall we think that the present paper can open the path for further experimental work on revealed intentions. One may, for instance, consider that not only the possibility of gains and losses but their magnitude would have an influence on social preferences. Building on the present approach and on the work by Fisman et al. (2007), further research might extend the findings presented here by investigating a richer set of revealed intentions using more complex choice sets than our, purposely simple, lines and points.

[14] We thank a reviewer for pointing out this connection to us.

References

Andreoni, J., & Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, *77*(5), 1607–1636.

Andreoni, J., & Miller, J. (2002). Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica*, *70*(2), 737–753.

Ashraf, N., Bohnet, I., & Piankov, N. (2006). Decomposing trust and trustworthiness. *Experimental Economics*, *9*(3), 193–208.

Bardsley, N. (2008). Dictator game giving: Altruism or artefact? *Experimental Economics*, *11*(2), 122–133.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122–142.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, *90*(1), 166–193.

Brandts, J., & Charness, G. (2011). The strategy versus the direct-response method: A first survey of experimental comparisons. *Experimental Economics*, *14*(3), 375–398.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, *117*(3), 817–869.

Chaudhuri, A., & Gangadharan, L. (2007). An experimental analysis of trust and trustworthiness. *Southern Economic Journal*, *73*(4), 959–985.

Cox, J. C. (2004). How to identify trust and reciprocity. *Games and Economic Behavior*, *46*(2), 260–281.

Cox, J. C., & Sadiraj, V. (2007). On modeling voluntary contributions to public goods. *Public Finance Review*, *35*(2), 311–332.

Cox, J. C., & Sadiraj, V. (2012). Direct tests of individual preferences for efficiency and equity. *Economic Inquiry*, *50*(4), 920–931.

Cox, J. C., Friedman, D., & Gjerstad, S. (2007). A tractable model of reciprocity and fairness. *Games and Economic Behavior*, *59*(1), 17–45.

Cox, J. C., Sadiraj, K., & Sadiraj, V. (2008a). Implications of trust, fear, and reciprocity for modeling economic behavior. *Experimental Economics*, *11*(1), 1–24.

Cox, J. C., Friedman, D., & Sadiraj, V. (2008b). Revealed altruism. *Econometrica*, *76*(1), 31–69.

Cox, J. C., Kerschbamer, R., & Neururer, D. (2016). What is trustworthiness and what drives it? *Games and Economic Behavior*, *98*, 197–218.

Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, *47*(2), 268–298.

Engelmann, D., & Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review*, *94*(4), 857–869.

Falk, A., & Kosfeld, M. (2006). The hidden costs of control. *The American Economic Review*, *96*(5), 1611–1630.

Fehr, E., & Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, *14*(3), 159–181.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, *114*(3), 817–868.

Fehr, E., Naef, M., & Schmidt, K. M. (2006). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment. *The American Economic Review*, *96*(5), 1912–1917.

Fisman, R., Kariv, S., & Markovits, D. (2007). Individual preferences for giving. *The American Economic Review*, *97*(5), 1858–1876.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, *1*(1), 1–12.

Krupka, E. L., & Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, *11*(3), 495–524.

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics*, *1*(3), 593–622.

List, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy*, *115*(3), 482–493.

McCabe, K. A., Rigdon, M. L., & Smith, V. L. (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior & Organization*, *52*(2), 267–275.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American Economic Review*, *83*(5), 1281–1302.

Train, K. (2009). *Discrete choice methods with simulation*. Cambridge: Cambridge University Press.

Wilcox, N. T. (2008). Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison. *Research in Experimental Economics*, *12*, 197–292.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, *13*(1), 75–98.